Automatic Chord Recognition from Audio

Alex Ray Emmons

Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA

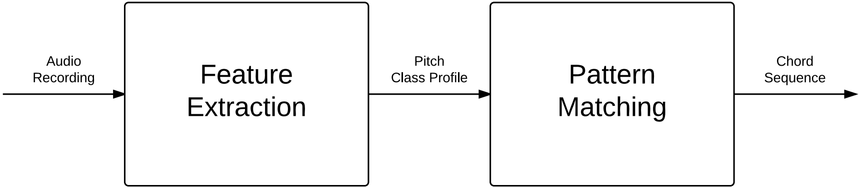
December 6, 2014

The Big Picture

- Used by researchers in the area of Music Information Retrieval (MIR) for tasks such as key detection, genre classification, and lyric interpretation.
- **Problem:** Performing chord analysis from audio by hand is time consuming and prone to error.
- **Potential Solution:** Automatic chord recognition systems.
- **Issues:** Noise in recordings, determining where chords change, complex music.

Solution

- Feature Extraction: Audio signals are processed to extract harmonic information, represented using a Pitch Class Profile.
- Pattern Matching: Chord labels are applied by matching chord models to the features that are present in the audio.
- Models can be generated either by hand or stochastically.



Outline

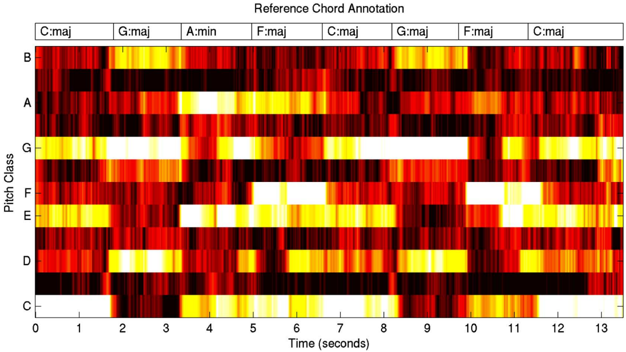
Outline

Preprocessing

- Preprocessing is an optimization step, performed during feature extraction, before a Pitch Class Profile (PCP) is generated.
- The goal of preprocessing is to remove as much background noise as possible from the audio file in an effort to provide a smooth and clear PCP.
- Two issues, background noise and overtones, usually addressed separately.

Pitch Class Profile

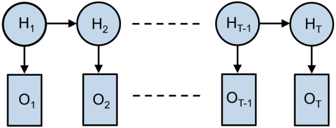
- Pitch Class Profile (PCP) measures energy in the 12 frequency regions where musical notes occur.
- Each row represents a pitch class, or note, and each column represents a frame, or period of time.
- Actual chord progression is shown above for reference.



Outline

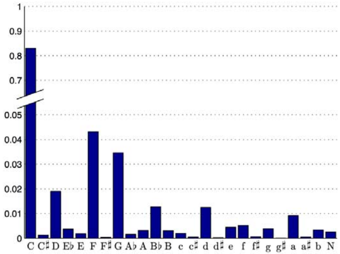
Hidden Markov Models

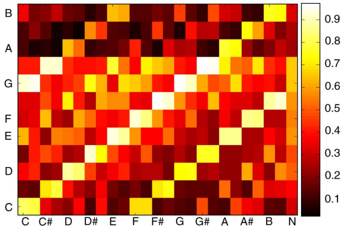
- A Hidden Markov Model (HMM) describes a sequence of states and transition probabilities.
- Transition probabilities are learned from labeled training data.
- The chords for the testing data are unknown, or hidden states. The PCP frames are the observed states.
- Observed states and transition probabilities are used to find the most likely sequence and eliminate unlikely transitions.



Hidden Markov Models

- Transition probabilities are learned by dividing the transitions to each chord by the total transitions from that chord.
- This is done for each chord, assigning a probability to all possible transitions.





Gaussian Mixture Models

- More detailed models for each chord are created by averaging features from multiple PCPs.
- Multiple variations of the same chord are represented using multiple Gaussian components.
- Like HMM, labeled training data is used and transition probabilities are learned for each chord.
- The observed chord is matched with each chord model to find the best fit.

Support Vector Machines

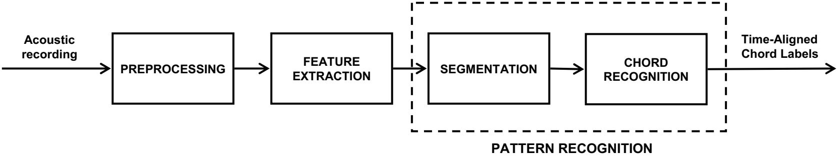
- Another type of supervised learning system.
- Trained with labeled testing data, chord labels are applied to segments of the test data.
- Only works on the kind of data it is trained on.
- Training procedure is complex for large datasets.

Outline

1. Feature Extraction
2. Pattern Matching
3. Research Cases
   - Case 1: Effects of Proper Signal Processing
   - Case 2: HMM Trained with Audio from Symbolic Data
   - Case 3: Importance of Individual Components
4. Conclusions

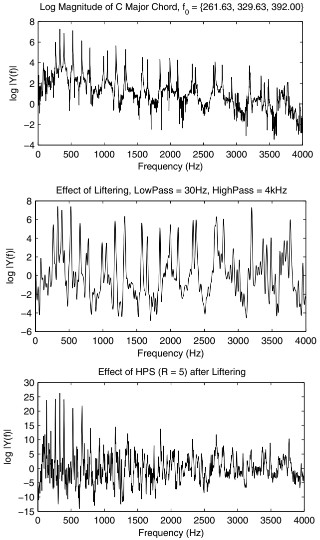
Case 1: Effects of Proper Signal Processing

- Two methods of preprocessing were used.
- Four feature vectors were compared.
- Two datasets were used.
- GMMs and SVMs are compared.



Preprocessing

- Homomorphic Liftering: Finds strong peaks in the frequency areas where notes occur.
- Harmonic Product Spectrum (HPS): De-emphasizes overtones, emphasizes chord tones.



Feature Extraction

- Four Feature Vectors (FV), or combinations of methods used for feature extraction were compared.
- Sample Rate is the audio resolution and Fast Fourier Transform (FFT) determines the resolution in the frequency domain.

	Type	**Sample Rate**	**FFT Length**	**Liftering**	**HPS Ratio**
FV1	FB	44100	32768	yes	5
FV2	PCP	11025	4096	no	1
FV3	PCP	44100	32768	no	1
FV4	PCP	44100	32768	yes	5

Datasets

- **Isolated chord dataset:** 7790 chords synthesized from Musical Instrument Digital Interface (MIDI) data on piano and strings.
- 80% used for training, 20% for testing. 3 complexity levels were tested.
- **Continuous single-instrument audio:** 50 hymns from the Trinity Hymnal, a MIDI collection of 761 hymns.
- 40 used for training, 10 for testing. FV4 and DS3 are used with an SVM.

	Label given in:		
Label	**DS1**	**DS2**	**DS3**
Major	Major	Major	Major
Minor	Minor	Minor	Minor
Major 7	-	Major	Major 7
Minor 7	-	Minor	Minor 7
Dom. 7	-	Major	Dom. 7
Dim.	Dim.	Dim.	Dim.
Full Dim.	-	Dim.	Full Dim.
Half Dim.	-	Dim.	Half Dim.
Augmented	Aug.	Aug.	Augmented
Sus. 4	-	-	Sus. 4
7 Sus. 4	-	-	7 Sus. 4

Results

Feature Vector	DS1	DS2	DS3
FV1	83.68	61.85	57.24
FV2	90.33	82.44	82.26
FV3	**91.76**	**84.20**	**84.09**
FV4	85.64	79.40	78.93

Table : Isolated chord recognition accuracy using GMM, training set: piano, testing set: piano

Feature Vector	DS1	DS2	DS3
FV1	68.06	42.00	33.62
FV2	42.72	18.60	16.30
FV3	43.49	22.00	18.31
FV4	**86.94**	**80.23**	**80.18**

Table : Isolated chord recognition accuracy using GMM, training set: piano, testing set: strings

Results

Feature Vector	DS1	DS2	DS3
FV1	93.43	88.21	86.52
FV2	94.78	93.26	93.13
FV3	**95.23**	**94.31**	**94.24**
FV4	90.56	88.08	87.74

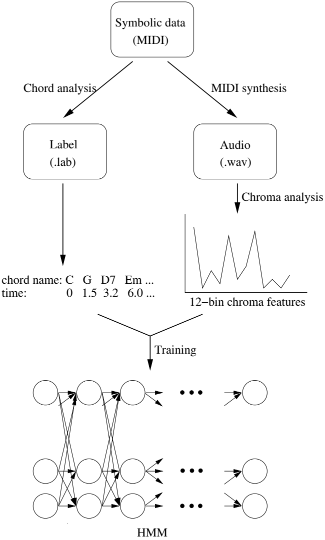
Table : Isolated chord recognition accuracy using SVM, training set: piano, testing set: piano

Number of Scatter Points			
3	**5**	**7**	**9**
72.73	87.77	88.07	**88.42**

Table : Continuous single-instrument recognition accuracy using FV4 and DS3, with varying number of scatter points

Case 2: HMM Trained with Audio-From-Symbolic Data

- Chord label data and audio files are created from the same MIDI data.
- Pitch Class profile is generated from the audio.
- These pieces are used to train a supervised HMM.



Datasets

- Two training datasets: 81 solo piano pieces, 196 string quartets by J.S. Bach, Beethoven, Mozart, and Haydn.
- Two testing datasets: 5 solo piano pieces, 5 string quartets selected from the Kostka and Payne's book.
- All combinations of training and testing data were tried.

Results

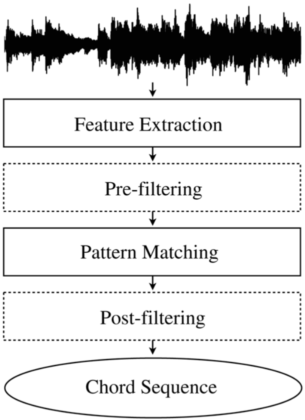
Training Data	Test Data	Recognition Rate
Piano	Piano	68.69
String Quartet	Piano	73.40
Piano & Strings	Piano	74.41
Piano	String Quartet	79.35
String Quartet	String Quartet	79.76
Piano & Strings	String Quartet	**80.16**

Table : Recognition results for all six possible training - test pairs in research case 2

Case 3: Importance of Individual Components

Four experiments:

- Using different combinations of preprocessing techniques during feature extraction.
- Pre-filtering using moving average filters, which look for noisy frames and smooth them across neighboring frames.
- Post-filtering using an HMM.
- Using both pre-filtering and post-filtering.



Datasets

- Each experiment was performed on 495 chord labeled songs.
- 180 Beatles songs, 20 Queen songs, 100 songs from the Real World Computing (RWC) pop dataset, and 195 songs from the US-Pop dataset.
- 5 groups of 99 songs were selected randomly, with four used for training and one for testing.

Results

- Each experiment was tried using 1, 5, 10, and 25 Gaussian components, with the best result shown.
- Increasing the number of components helped when using HMMs because they are dependent on transition probabilities.
- For preprocessing, high and low frequencies were de-emphasized and log compression was used, which limits dynamic range caused by different instruments and volumes.

Expt.	Highest Accuracy	Pre-filtering	Pattern Matching	Post-filtering
1	58.30	-	1 Gaussian component	-
2	71.22	Moving average filters	1 Gaussian component	-
3	**77.90**	-	25 Gaussian components	HMM
4	77.58	Moving average filters	25 Gaussian components	HMM

Table : Results from research case 3, showing the highest accuracy in each experiment, and the components used to achieve it

Outline

Conclusions

- Isolated chords were more of a proof of concept.
- Highest accuracy is around 88%, achieved using homomorphic liftering, HPS, chord segmentation, and SVMs for labeling.
- This system was dependent on the training data, and SVMs don't work as well for large datasets.
- Systems can be specifically tailored to type types of instruments and chords present in the dataset.
- Advances are being made in more general models that can provide chords for any given song.

Thanks!

Thank you for your time and attention!

Contact:

- `emmon046@morris.umn.edu`

Questions?

References I

Morman, Joshua and Rabiner, Lawrence.
A System for the Automatic Segmentation and Classification of Chord Sequences.
In Proceedings of the 1st ACM workshop on Audio and music computing multimedia (AMCMM '06). ACM, New York, NY, USA, 1-10. DOI=10.1145/1178723.1178725

Lee, Kyogu and Slaney, Malcolm
Automatic Chord Recognition from Audio Using a Supervised HMM Trained with Audio-from-symbolic Data.
In Proceedings of the 1st ACM workshop on Audio and music computing multimedia (AMCMM '06). ACM, New York, NY, USA, 11-20. DOI=10.1145/1178723.1178726

References II

TaeMin Cho and Bello, J.P.
On the Relative Importance of Individual Components of Chord Recognition Systems.
IEEE/ACM Trans. Audio, Speech and Lang. Proc. 22, 2 (February 2014), 477-492. DOI=10.1109/TASLP.2013.2295926

McVicar, M. and Santos-Rodriguez, R. and Yizhao Ni and Tijl De Bie
Automatic Chord Estimation from Audio: A Review of the State of the Art
IEEE/ACM Trans. Audio, Speech and Lang. Proc. 22, 2 (February 2014), 556-575. DOI=10.1109/TASLP.2013.2294580